A System for Creating Virtual Reality Content from Make-Believe Games

Adela Barbulescu[1], Maxime Garcia[1], Antoine Begault[1], Laurence Boissieux[1], Marie Paule Cani[1], Maxime Portaz[2], Alexis Viand[1], Romain Dulery[1], Pierre Heinish[1], Remi Ronfard[1] and Dominique Vaufreydaz[2]

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK, 38000 Grenoble, France

[2] Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LIG, 38000 Grenoble, France

Author version

Abstract

Pretend play is a storytelling technique, naturally used from very young ages, which relies on object substitution to represent the characters of the imagined story. We propose a system which assists the storyteller by generating a virtualized story from a recorded dialogue performed with 3D printed figurines. We capture the gestures and facial expressions of the storyteller using Kinect cameras and IMU sensors and transfer them to their virtual counterparts in the story-world. As a proof-of-concept, we demonstrate our system with an improvised story involving a prince and a witch, which was successfully recorded and transferred into 3D animation.

1 Introduction

A new challenge in virtual environments is the introduction of storytelling, with increasing interest in providing virtual storytelling tools that can be used to teach narrative skills to young children [1, 3]. While early work has investigated the use of digital puppets [5] and interactive spaces [2], many researchers have noted that tangible interaction with actual physical puppets is more engaging. Indeed real-time interaction with a magic mirror metaphor - showing the story in the virtual world as it is being played in the real world - causes problems of divided attention between those two mental work spaces. In contrast, we describe a system where the storyteller is allowed to improvise freely and later to re-watch the imagined story. This system helps in easily creating content for virtual environments, even by children.

Our system allows the storyteller to improvise a scene with two figurines in hand by alternatively interpreting the lines of two characters in front of two Kinect cameras (see Figure 1). The front camera records the storyteller's facial expressions and head movements. The top camera records the movements of the figurines, which are also equipped with inertial measurement units (IMU). The free improvisation of the storyteller is used to animate the two characters: the head movements and facial expressions of the storyteller are transferred to the corresponding character's head, while the figurine motions are transferred to the bodies of the characters.

2 Recording system

Recent storytelling systems which use RGB-D sensors [4, 6] are sensitive to occlusion. We reduce this problem by combining the top Kinect with IMU sensors which are placed inside the figurines. This configuration was set to address the tasks of puppet identification and localization. In addition, the IMU sensors provide the angular positions which allow the recovery of the 6D path for each figurine. We also enrich the recordings with body pose, facial expression and voice of the storytellers. The front kinect records the storyteller's face features using the FaceShift software. This markerless motion capture system returns accurate head rotation and translation, gaze direction and facial expressions.

*Institute of Engineering Univ. Grenoble Alpes

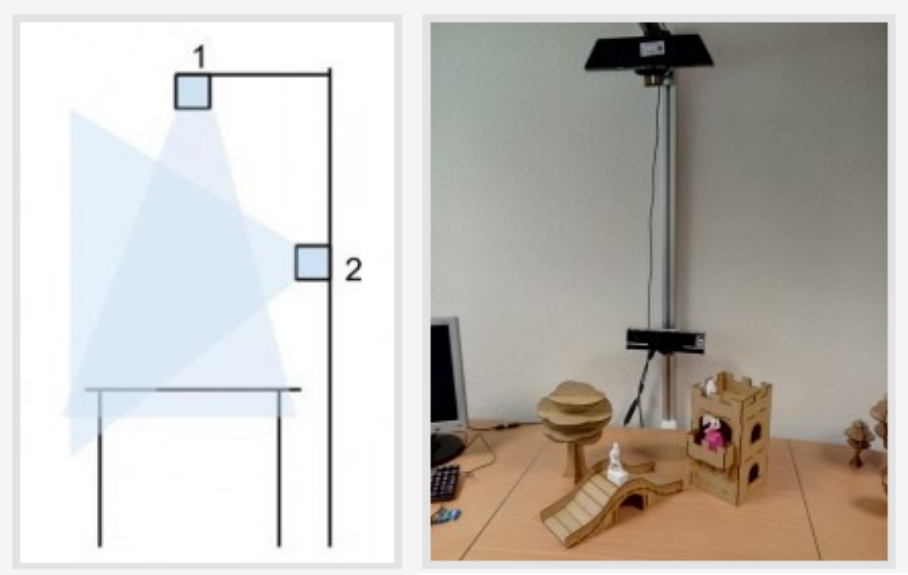

Figure 1: Acquisition setup : one Kinect is looking down to follow figurines (1), another is following narrators (2).

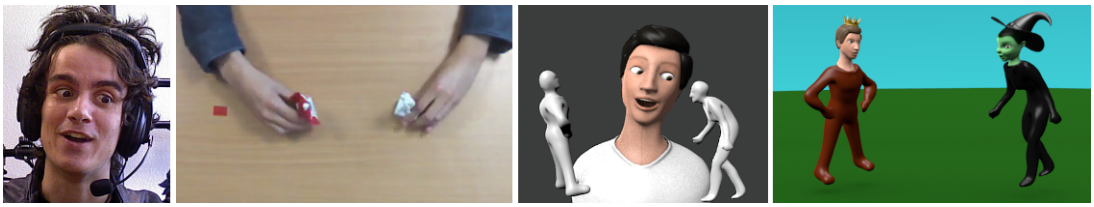

Figure 2: Corresponding frame for: (a) front Kinect video data displaying the storyteller's expressions, (b) top Kinect video data displaying the figurines, (c) a reconstructed scene including a virtual storyteller and virtual figurines (d) the final animated version of the story.

3 Animation Layers

We propose several layers of animation for the movements of the virtual character: (1) rigid body motion (rotation and translation) is transferred directly from the tracked movements of the corresponding figurine, (2) head rotation, gaze, facial expressions and voice are transferred directly from the storyteller if the corresponding figurine is interpreted, or are automatically generated otherwise, (3) body and head rotations and facial expressions may be re-adapted.

4 Story Analysis

The storyteller can only interpret one character at a time. Determining which figurine is being interpreted at each moment is an essential task. For this we first extract sentences and then assign them to the corresponding figurines i.e. the ones that are focused during speech. We recorded a training performance, in which a female storyteller interprets a male and a female character and we computed a set of parameters as an optimization problem for correctly assigning a figurine with the corresponding sentences. These parameters can then be used for a new storytelling performance.

Voice Analysis. A strategy in impersonating characters is changing the voice pitch, intensity or rhythm. Because our tests showed that the storytellers have difficulty maintaining distinctive pitch strategies for the characters, we only use the intensity of the voice signal to extract sentences. The sentences are extracted by separating silent frames from speech frames using an intensity threshold of 50 dB, and then by concatenating the successive speech frames.

Gaze Analysis. Gaze direction is an important cue because storytellers tend to look at the figurine that they are currently interpreting. Therefore, the focused figurine is assigned by determining whether the gaze is oriented towards the left or the right relative to an imaginary vertical plane which passes through the center of the storyteller and equally divides the space between the figurines.

Movement Analysis. Storytellers tend to move more the focused figurine. In order to determine the focused figurine during speech, we compare the amount of motion variation for a set of frames for the two figurines. We compute the sum of 1-norm between the current position and orientation and the one at a previous frame, for the last N frames. In our experiment, the N number which obtains best assignment rate is 78, for a rate of 30 fps.

Focus Choice. The focused figurine is chosen by solving an optimization problem where we assigned weight coefficients for the two methods. The best result is obtained when the weight attributed to the movement analysis method is 1 and for the gaze analysis is 0, showing that the figurine movement is a better cue for assigning the focused character at the level of the sentence.

Faceshift Motion Adaptation. Once the focus is obtained, we directly transfer the expressions and head motion of the storyteller to the assigned characters. Next, we perform adaptations, such as head pitch scaling and eyelid raising, in order to correct the storyteller's motions which are caused by manipulating the figurines: head and gaze are oriented downwards. We also introduce modifications for satisfying cinematographic rules in the virtual scene: we rotate the bodies with 25 degrees towards the camera, such that the characters' faces are visible and maintaining a dialogue impression.

Splitting/Interpolation. We split the adapted motion data according to focus intervals. For the non-focus intervals we opt for neutral facial expressions, such that the motion looks natural and does not distract attention from the speaking character. Head motion in non-focused intervals is obtained using linear interpolation.

5 A Pilot Study

In order to evaluate our system, we recorded a male storyteller delivering a short story with two characters: the prince and the witch. The entire story lasts 1 minute and 25 seconds and consists of 12 sentences alternating between the two characters. Figure 2 illustrates steps in the generation of the animated story.

The voice analysis algorithm extracts sentences with a total of 84% speech frames correctly identified. Using the gaze analysis algorithm leads to correctly assigning 85% of the speech frames to the focused characters, while using the movement analysis obtains 89% recognition rate. The focus choice algorithm leads to a final correct assigning of 89% speech frames since we only use the figurine movement to choose the focused figurines. We notice that gaze analysis is less reliable than figurine movement because the storyteller tends to switch the gaze direction to the other figurine before the sentence ends.

6 Discussion

We presented a system for the virtualization of pretend-play for children which allows a storyteller to manipulate figurines while interpreting the imagined characters. Our experiments indicate that a combination of cues (voice prosody, figurine motion, storyteller gaze) is necessary for determining speaking turns between the two characters. This result needs to be confirmed with more extensive testing involving children.

This system represents a tool for easily creating virtual reality content for inexperienced users, especially children. In future work we would like to enhance this system by allowing direct interactions inside the virtual reality generated content. The system's usability for immersive interactive storytelling can be evaluated in a similar manner to [7]. We will also look at inferring the gestures intended by the storyteller (walking, jumping etc) and transferring them to the virtual characters.

7 Acknowledgement

Prototyping was done using Amiqual4Home facilities (ANR-11-EQPX-0002). The work was supported by the European Research Council advanced grant EXPRESSIVE (ERC-2011-ADG 20110209) and the PERSYVAL-Lab (ANR-11-LABX-0025-01) Labex.

References

[1] S. Benford, B. Bederson, K. Åkesson, V. Bayon, A. Druin, P. Hansson, J. Hourcade, R. Ingram, H. Neale, C. O'Malley, and Others. Designing storytelling technologies to encouraging collaboration between young children. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2(1):556–563, 2000.

[2] A. F. Bobick, S. S. Intille, J. W. Davis, F. Baird, C. S. Pinhanez, L. W. Campbell, Y. a. Ivanov, A. Schütte, and A. Wilson. The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment. *Presence: Teleoperators and Virtual Environments*, 1999.

[3] F. Garzotto, P. Paolini, and A. Sabiescu. Interactive storytelling for children. *Proceedings of the 9th International Conference on Interaction Design and Children IDC 10*, 2(1):356, 2010.

[4] S. Gupta, S. Jang, and K. Ramani. Puppetx: A framework for gestural interactions with user constructed playthings. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, pages 73–80, New York, NY, USA, 2014. ACM.

[5] B. Hayes-Roth and R. van Gent. Story-Making with Improvisational Puppets. In *Proceedings of the first international conference on Autonomous agents (AGENTS '97)*, pages 1–7, 1997.

[6] R. Held, a. Gupta, B. Curless, and M. Agrawala. 3d puppetry: A kinect-based interface for 3d animation. *Proceedings of UIST*, 2012.

[7] J.-L. Lugrin, M. Cavazza, D. Pizzi, T. Vogt, and E. André. Exploring the usability of immersive interactive storytelling. In *Proceedings of the 17th ACM symposium on virtual reality software and technology*, pages 103–110. ACM, 2010.